

NIL-UCM at MER-TRANS 2026: Integrating Linguistic Knowledge into Prompting Strategies for Spanish Text Simplification

Alberto Díaz¹, Daniel Fernández¹

¹Facultad de Informática and ITC, Universidad Complutense de Madrid, Spain

Abstract

This paper describes the participation of the NIL-UCM team in the MER-TRANS (Multilingual Easy-to-Read Translation) shared task. Building on our previous work in the CLEARS challenge, we apply a methodology based on explicit linguistic instructions within prompts using a Large Language Model (Mistral-7B-Instruct-v0.3). We evaluate our system on the Spanish portion of the iDEM corpus, which focuses on texts related to democratic participation. Our results suggest that while structured prompts provide a solid foundation for simplification, the specific requirements of the democratic participation domain may favor more flexible prompting strategies.

Keywords

automatic simplification, easy-to-read, linguistic knowledge, prompting

1. Introduction

This paper presents the participation of the NIL-UCM team in the MER-TRANS challenge [1], held as part of IberLEF [2]. The MER-TRANS shared task aims to advance automatic methods for producing easy-to-read (E2R) versions of texts. Accessible language is vital for supporting the participation of individuals with reading comprehension difficulties, aligning with global inclusion goals and standards such as UNE 153101 EX standard [3]. Our participation is based on adapting a proposal that leverages a complex prompt following the UNE standard, which was previously used in the CLEARS (IberLEF 2025) task [4, 5].

2. Task Description

The objective of MER-TRANS is to automatically transform complex excerpts into simplified, readable versions that preserve the original meaning. Texts come from a domain-focused corpus (democratic participation). They have been simplified by experts following easy-to-read recommendations and validation procedures.

2.1. Dataset

The task uses the iDEM corpus, which contains original and simplified text pairs related to democratic participatory processes. Unlike news-centric datasets used in previous challenges, iDEM covers informative, policy, and news content specifically tailored to civic engagement. There are data in different languages: Catalan, Italian, Spanish and Arabic [6, 7].

There is no specific training set released for the task. External resources can be used to solve the task. As trial data there is a full document per language together with its easy-language adaptation.

To better understand the nature of the original texts and those provided to participants as references for adaptation, we include a fragment from the Spanish trial text in its two versions:

IberLEF 2026, September 2026, León, Spain

✉ adiazest@ucm.es (A. Díaz); dferna17@ucm.es (D. Fernández)

ORCID 0000-0003-1966-3421 (A. Díaz)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Original

"Abogados o letrados

Los abogados o letrados son aquellos profesionales encargados de la defensa y dirección de las partes en los procedimientos judiciales y administrativos, asumiendo también con carácter amplio labores de asesoramiento y consejo jurídico en el ámbito extrajudicial para evitar conflictos."

Easy-to-Read

"Debes contactar con abogados o letrados.

Los abogados o letrados son los profesionales encargados de defender y dirigir los procesos judiciales y administrativos. Ellos también pueden asesorarte, orientarte y dar consejos para evitar conflictos fuera del proceso judicial."

As for the test files, they contain the excerpts that must be adapted automatically. These texts belong to the same domain but they differ from the text used as trial data. There are 16 documents in the test set, with 4163 excerpts.

2.2. Evaluation

MER-TRANS employs a multidimensional evaluation suite to capture different facets of simplification:

- SARI: simplification-focused. It measures add/keep/delete operations vs inputs and references.
- BLEU: surface similarity. It compares system output to reference simplifications.
- BERTScores and MeaningBERT: semantic similarity. They measure signals of meaning preservation.
- Readability/Complexity. It may complement the core metrics to assess accessibility.

The official evaluator is available in a Git repository,

3. Methodology

The methodology used is based on that presented by the NIL-UCM team in the CLEARs (IberLEF 2025) task. The idea is to include in a prompt linguistic knowledge to perform the simplification in a better way.

To carry out the task of automatic text simplification, we employed Mistral-7B-Instruct-v0.3, a large language model (LLM) with 7 billion parameters, designed for automatic text generation. This model is a fine-tuned version of the base model Mistral-7B-v0.3, specifically trained to follow human instructions. This fine-tuning process, known as instruct fine-tuning, enables the model to respond more helpfully, coherently, and in a task-oriented manner. Thanks to its ability to interpret and execute instructions, the model is particularly suitable for tasks such as text simplification, which require transforming content while preserving its original meaning but reducing its linguistic complexity¹.

Mistral-7B-Instruct-v0.3 met all the requirements established by our methodology:

1. A model specialized in coherent and fluent text generation.
2. No need for additional fine-tuning with the competition's training data.
3. Ability to follow precise linguistic instructions.
4. Significantly lower cost compared to models from OpenAI, Meta, and Google.

¹The model description is available at <https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3>

3.1. Prompt design

In order to integrate this linguistic knowledge into our model, the most relevant guidelines and recommendations on the subject have been compiled. Our main reference has been the experimental standard UNE 153101:2018 EX [3]. This standard, which is the primary reference in Spain, includes both mandatory guidelines and recommendations, often illustrated with incorrect and correct examples. García Muñoz [8] is also highly relevant, as he systematically presents drafting and evaluation proposals based on previous experiences. Although both documents also include proposals related to design and layout, we have focused on those related to orthotypography, lexis, morphosyntax, style, and the organization of information.

The prompt used was designed in accordance with the guidelines and recommendations of Easy-to-Read outlined previously. It was the prompt used in the Clear task in Iberlef 2026 [4, 5]. The prompt is divided into eight parts, such that general instructions are followed by seven sections grouping a substantial number of guidelines along with their examples: general objective, punctuation marks and symbols, line formatting, lexis, numbers, morphosyntax, and information structure. Our prompt aimed to reflect the higher level of precision and thoroughness characteristic of Easy-to-Read, as evidenced by its length of 375 lines.

It is important to remark that this prompt was designed for document simplification, not for sentence simplification. However, we thought that it could work in the excerpt simplification scenario in MER-TRANS 2026. The main aim of our participation is to demonstrate this hypothesis.

We have submitted two configurations for the Spanish track:

- RUN 1: mistral with prompt containing the explicit Easy-to-Read guidelines described before (the same prompt used in Clears 2025).
- RUN 2: A variant described as "Easy-to-Read without organization/restrictions". This version aimed to test if a less constrained prompt would allow the model to better adapt to the sentence simplification scenario.

4. Results

The official results for the NIL-UCM team in the Spanish track are summarized in Table 1:

Table 1

Results obtained by the different evaluated methods.

Method	BLEU-Orig	BLEU-Gold	SARI	BERT-Score
RUN 2 (No restrictions)	11.3818	4.7980	40.5175	0.9026
RUN 1 (Guidelines)	6.0574	2.5775	39.6873	0.8780

RUN 2 consistently outperformed RUN 1 across all metrics. This suggests that the elimination of organization guidelines improves simplification.

As can be seen in Table 2, the results obtained by other systems show that there is significant room for improvement in our developed system. One of the main issues could be that the system developed for the CLEARS task was focused on document simplification, even though the rules specified in the prompt are mostly at the sentence level. The other problem is related to the hallucinations generated by the system, which are likely due to the excessive complexity of the prompt. However, despite the low BLEU-Gold scores, our BERTScore of 0.9026 indicates high semantic fidelity to the original content.

5. Conclusions and Future Work

Our participation in MER-TRANS suggests that LLMs are powerful tools for cognitive accessibility, but they require an adjustment in the granularity of instructions. The superiority of RUN 2 suggests that,

Table 2

System results in the simplification task.

Team	Language	Method	BLEU-Orig	BLEU-Gold	SARI	BERT-Score
ClearText	ES	RUN1	18.0013	15.4483	47.0168	0.9220
VICOMTECH	ES	RUN3	20.7912	14.1086	45.5843	0.9241
HULAT1	ES	RUN2	33.1815	17.7135	45.0182	0.9272
HULAT1	ES	RUN3	31.5608	17.1627	44.7465	0.9272
VICOMTECH	ES	RUN1	12.4857	9.3316	44.1880	0.9114
HULAT2	ES	RUN1	35.5730	17.5715	44.0543	0.9263
HULAT1	ES	RUN1	37.1581	18.2081	43.3874	0.9276
BASELINE1	ES	RUN1	18.9491	10.8850	43.2873	0.9130
HULAT2	ES	RUN2	34.1564	16.1484	43.1049	0.9248
VICOMTECH	ES	RUN2	32.0715	14.7485	42.1724	0.9246
HumanAI-UCM	ES	RUN2	26.9792	11.8405	41.6196	0.9111
FACILE	ES	RUN1	3.0009	2.8644	41.6129	0.8783
NIL_UCM	ES	RUN2	11.3818	4.7980	40.5175	0.9026
HumanAI-UCM	ES	RUN1	38.3402	15.0830	40.4163	0.9209
NIL_UCM	ES	RUN1	6.0574	2.5775	39.6873	0.8780
HULAT2	ES	RUN3	2.4037	5.3373	38.5136	0.9114
BASELINE2	ES	RUN1	29.0913	9.3904	37.1761	0.9019
HumanAI-UCM	ES	RUN3	45.0236	14.5298	36.2450	0.9125
DoNothing	ES	RUN1	100.0000	23.2176	13.8687	0.9297

in specific domains such as the iDEM corpus, fewer restrictions may favor better language adaptation. Future work will focus on reducing the prompt’s cognitive load to minimize hallucinations and exploring few-shot techniques specific to the democratic domain.

Declaration on Generative AI

During the preparation of this work, the authors used Gemini in order to check grammar and spelling, as well as to paraphrase and reword text. After using this tool, the authors reviewed and edited the content as needed, and they take full responsibility for the publication’s content.

6. Acknowledgements

This publication is part of the R&D&I project HumanAI-UI, Grant PID2023-148577OB-C22 (Human-Centered AI: User-Driven Adaptative Interfaces-HumanAI-UI) funded by MICI-U/AEI/10.13039/501100011033 and by FEDER/UE.

References

- [1] H. Saggion, M. Tareh, N. Khallaf, D. Adanza, S. Bott, N. P. Rojas, A. Rascón, S. Szasz, Overview of MER-TRANS at IberLEF 2026: First shared task on multilingual easy-to-read translation, *Procesamiento del Lenguaje Natural 77* (2026).
- [2] A. Bonet-Jover, J. Á. González-Barba, L. Chiruzzo, Overview of IberLEF 2026: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2026)*, co-located with the 42nd Conference of the Spanish Society for Natural Language Processing (SEPLN 2026), CEUR-WS.org, 2026.
- [3] AENOR, UNE 153101 EX. Lectura Fácil. Pautas y recomendaciones para la elaboración de documentos, AENOR, 2018.

- [4] D. Fernández, A. Díaz, Integrating linguistic knowledge into prompting strategies for spanish text simplification: Insights from the nil-ucm participation, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), volume 4098 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025. URL: <https://ceur-ws.org/Vol-4098/>.
- [5] D. Fernández, Estrategias de Prompting Basadas en Conocimiento Lingüístico para la Simplificación de Textos en Español, Trabajo de fin de máster, Facultad de Informática, Universidad Complutense de Madrid, Madrid, España, 2025. Tutorizado por Alberto Díaz.
- [6] S. Bott, V. Riegler, H. Saggion, A. Rascón-Alcaina, N. Khallaf, A multilingual human annotated corpus of original and easy-to-read texts to support access to democratic participatory processes, in: Proceedings of the Language Resources and Evaluation Conference (LREC 2026), European Language Resources Association (ELRA), Palma, Mallorca, Spain, 2026.
- [7] N. Khallaf, S. Sharoff, Automatic difficulty classification of arabic sentences, in: Proceedings of the Sixth Arabic Natural Language Processing Workshop, Association for Computational Linguistics, Kyiv, Ukraine (Virtual), 2021, pp. 105–114. URL: <https://aclanthology.org/2021.wanlp-1.12>. doi:10.18653/v1/2021.wanlp-1.12.
- [8] O. García Muñoz, Lectura fácil: métodos de redacción y evaluación, Real Patronato sobre Discapacidad, 2012.